JEI: Just-Enough-Information Paradigm for roduction Scheduling in a Manufacturing Supply Network

id D. Grossman, Peter M. Will, Mirza M. Beg and Victor D. Lee

19960731 048

96 7 18 519

JEI: Just-Enough-Information Paradigm for Production Scheduling in a Manufacturing Supply Network

David D. Grossman, Peter M. Will, Mirza M. Beg and Victor D. Lee

ISI/RR-95-402

May 1995

University of Southern California
Information Sciences Institute
4676 Admiralty Way, Marina del Rey, CA 90292-6695
310-822-1511

REPORT DOCUMENTATION PAGE

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE **May 1995**	3. REPORT TYPE AND DATES COVERED **Research**

4. TITLE AND SUBTITLE
JEI:Just-Enough-Information Paradigm for Production Scheduling in a Man-ufacturing Supply Network

5. FUNDING NUMBERS
DABT 63-92-C-0052

6. AUTHOR(S)
David D. Grossman

Peter M. Will

Mirza M. Beg and Victor D. Lee

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

USC INFORMATION SCIENCES INSTITUTE
4676 ADMIRALTY WAY
MARINA DEL REY, CA 90292-6695

8. PERFORMING ORGANIZATON REPORT NUMBER
ISI/RR-95-402

9. SPONSORING/MONITORING AGENCY NAMES(S) AND ADDRESS(ES)
ARPA ESTO
3701 N. Fairfax Drive
Arlington, VA 22203-1714

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
A shorter version of this paper is being submitted to a journal.

12A. DISTRIBUTION/AVAILABILITY STATEMENT
UNCLASSIFIED/UNLIMITED

12B. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

We describe an emerging problem in manufacturing and its potential solution. The problem deals with the optimization of manufacturing supply networks, as opposed to the optimization of individual company manufacturing facilities. The opportunity for cost savings is great, but the need to preserve confidentiality of each company's data presents a major impediment. Just-Enough-Information (JEI) is a new paradigm for production scheduling, based on exchange of minimal proprietary data among companies. We describe a specific new algorithm to implement JEI, based on exchange of Demand-Availability-Order (DAO) data. Evidence from simulation studies indicates that this methodology is highly promising. Practical application depends on generalizing DAO and on complement-ing it with a broad range of related systems, including real time coarse capacity management. The paper also addresses these comprehensive requirements

14. SUBJECT TERMS
Information, Inventory Control, JIT, MRP, Manufacturing, Production Information Control, Production Scheduling.

15. NUMBER OF PAGES
22

16. PRICE CODE

17. SECURITY CLASSIFICTION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UNLIMITED

GENERAL INSTRUCTIONS FOR COMPLETING SF 298

The Report Documentation Page (RDP) is used in announcing and cataloging reoprts. It is important that this information be consistent with the rest of the report, particularly the cover and title page. Instructions for filling in each block of the form follow. It is important to stay within the lines to meet optical scanning requirements.

Block 1. Agency Use Only (Leave blank).

Block 2. Report Date. Full publication date including day, month,a nd year, if available (e.g. 1 jan 88). Must cite at least the year.

Block 3. Type of Report and Dates Covered. State whether report is interim, final, etc. If applicable, enter inclusive report dates (e.g. 10 Jun 87 - 30 Jun 88).

Block 4. Title and Subtitle. A title is taken from the part of the report that provides the most meaningful and complete information. When a report is prepared in more than one volume, repeat the primary title, add volume number, and include subtitle for the specific volume. On classified documents enter the title classification in parentheses.

Block 5. Funding Numbers. To include contract and grant numbers; may include program element numbers(s), project number(s), task number(s), and work unit number(s). Use the following labels:

C	- Contract	PR	- Project
G	- Grant	TA	- Task
PE	- Program Element	WU	- Work Unit Accession No.

Block 6. Author(s). Name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. If editor or compiler, this should follow the name(s).

Block 7. Performing Organization Name(s) and Address(es). Self-explanatory.

Block 8. Performing Organization Report Number. Enter the unique alphanumeric report number(s) assigned by the organization performing the repor.

Block 9. Sponsoring/Monitoring Agency Names(s) and Address(es). Self-explanatory

Block 10. Sponsoring/Monitoring Agency Report Number. (If known)

Block 11. Supplementary Notes. Enter information not included elsewhere such as: Prepared in cooperation with...; Trans. of ...; To be published in... When a report is revised, include a statement whether the new report supersedes or supplements the older report.

Block 12a. Distribution/Availability Statement. Denotes public availability or limitations. Cite any availability to the public. Enter additional limitations or special markings in all capitals (e.g. NOFORN, REL, ITAR).

DOD	- See DoDD 5230.24, "Distribution Statements on Technical Documents."
DOE	- See authorities.
NASA	- See Handbook NHB 2200.2.
NTIS	- Leave blank.

Block 12b. Distribution Code.

DOD	- Leave blank.
DOE	- Enter DOE distribution categories from the Standard Distribution for Unclassified Scientific and Technical Reports.
NASA	- Leave blank.
NTIS	- Leave blank.

Block 13. Abstract. Include a brief (Maximum 200 words) factual summary of the most significant information contained in the report.

Block 14. Subject Terms. Keywords or phrases identifying major subjects in the report.

Block 15. Number of Pages. Enter the total number of pages.

Block 16. Price Code. Enter appropriate price code (NTIS only).

Blocks 17.-19. Security Classifications. Self-explanatory. Enter U.S. Security Classification in accordance with U.S. Security Regulations (i.e., UNCLASSIFIED). If form contins classified information, stamp classification on the top and bottom of the page.

Block 20. Limitation of Abstract. This block must be completed to assign a limitation to the abstract. Enter either UL (unlimited) or SAR (same as report). An entry in this block is necessary if the abstract is to be limited. If blank, the abstract is assumed to be unlimited.

JEI: Just-Enough-Information Paradigm for Production Scheduling in a Manufacturing Supply Network[1]

David D. Grossman, Peter M. Will, Mirza M. Beg and Victor D. Lee
University of Southern California/Information Sciences Institute
4676 Admiralty Way, Marina Del Rey, CA 90292. USA.
e-mail:grossman@isi.edu

Abstract

We describe an emerging problem in manufacturing and its potential solution. The problem deals with the optimization of manufacturing supply networks, as opposed to the optimization of individual company manufacturing facilities. The opportunity for cost savings is great, but the need to preserve confidentiality of each company's data presents a major impediment. Just-Enough-Information (JEI) is a new paradigm for production scheduling, based on exchange of minimal proprietary data among companies. We describe a specific new algorithm to implement JEI, based on exchange of Demand-Availability-Order (DAO) data. Evidence from simulation studies indicates that this methodology is highly promising. Practical application depends on generalizing DAO and on complementing it with a broad range of related systems, including real time coarse capacity management. The paper also addresses these comprehensive requirements[2].

1. Introduction

The core of operational control for manufacturing is production scheduling. Its task is to optimize some measure of performance, based on the state of the manufacturing system and a set of external customer demands. State changes are governed by internal interdependencies among inventories (raw materials, work-in-process, and finished goods), resources (equipment and labor), and funds. Uncertainty, planned change, and unplanned change make the production scheduling task much too complex for simple solution.

In pervasive practice, the highest level of production scheduling deals with the optimization of individual company manufacturing facilities. An emerging problem in manufacturing is the optimization of manufacturing *supply networks*. The opportunity for cost savings is great, but the need to preserve confidentiality of each company's data presents a major impediment.

In this paper, we propose a new Just-Enough-Information (JEI) paradigm for production scheduling, tailored to the optimization of manufacturing supply networks. Section 2 discusses the economic motivation for supply network optimization and explains how this problem relates to Electronic Commerce. Section 3 provides background information on the prior art in production scheduling paradigms and explains their limitations in the context of variances in a supply network. Section 4 introduces the JEI paradigm, focusing on its prescrip-

tion for exchange of specific non-proprietary production data among companies. We describe the form of this Demand-Availability-Order (DAO) data and how it would be processed. Section 5 presents Monte Carlo simulation studies of an abstract model of a supply network, comparing our approach to other approaches, and indicating that the DAO methodology is highly promising. Section 6 discusses a range of generalizations needed for eventual practical application of JEI. Section 7 discusses the need for any new supply network production scheduler to provide a means for coarse capacity management in real time, and proposes a specific approach to this problem. Section 8 gives a brief outline of the system architecture need to implement the new paradigm. The conclusions are summarized in Section 9.

2. Motivation

The importance of production scheduling is evidenced by the fact that a major fraction of manufacturing executives' time is devoted to handling exceptions (mainly due to parts) and dealing with their associated costs. Improvements in production scheduling can reduce the cost of indirect labor associated with production. These overhead costs include every keystroke to query the production information system, every phone call to obtain on-time delivery of parts and expedite

2. On-line information may be accessed at URL: http://irobot.isi.edu/flex/flexmain.html.

orders, every intervention to get production running after a stoppage, and almost every transaction in the purchasing department. Collectively, these costs constitute most of manufacturing overhead [16].

At a national level, the monetary implications of production scheduling are immense. For instance, electronic equipment manufacturing in the US is roughly a $160 billion industry. The typical manufacturer spends about 55% of revenue on manufacturing cost, and 40% of manufacturing cost is overhead. Thus, overhead for electronic equipment manufacturing in the US amounts to about $35 billion per year. Additionally, these same manufacturers maintain inventories of roughly $25 billion (not including another $10 billion in components)[1]. Similar huge numbers apply in many other industry sectors [2], [3]. The potential improvement in these numbers is the motivation for US support of the development of improved policies for production scheduling.

From the point of view of individual manufacturers, the motivation is even higher. The final assembler of electronic equipment spends 50% of his manufacturing cost on materials. Thus, in addition to the 40% overhead he incurs directly, he knows he is incurring 40% overhead hidden within his 50% materials cost, and so on recursively up the supply network. The overall fraction of his manufacturing cost due to overhead somewhere in the network is much larger than 40%.

The data required by a new paradigm is already largely on-line. Essentially all but very small electronics equipment manufacturers use Production Information Control Systems (PICS) to monitor the state of their factories. Within these PICS are modules to support various functions, e.g., Sourcing, Purchasing, Order Tracking, Shipping and Receiving, Billing and Accounts Receivable, Finance, Quality Assurance, Equipment Engineering, Inventory Control, Capacity Planning, Production Planning, and Master Production Scheduling. Each PICS may be attached to a real-time Plant Floor System that sends control signals to equipment and monitors activity.

What is lacking is a means to interconnect the PICS across a supply network and a means to use that interconnection intelligently. Today, nearly all inter-company activities are outside the scope of PICS. For example, the purchasing function depends on an agent at one company phoning or faxing his counterpart at another company to get quotes (specifications, price, quantity, and availability) to place orders, and to inquire about the state of pending orders.

During the past decade, the manufacturing environment has changed because of the pervasive application of computers and networking. The new environment is characterized by worldwide supply networks, worldwide markets, rapidly changing products, and rapidly changing demand. The advent of Electronic Data Interchange (EDI) has made it possible for data to be exchanged among companies via computer networks, and the trend towards Electronic Commerce (EC) [7] will make the use of EDI more prevalent. But EC does

---

1. In the domain of military equipment, the Defense Logistics Agency has an inventory of roughly $100 billion, which it would like to reduce to $6 billion by the year 1998.

not understand production data. As a result, the first phase of the transition to EC will simply replace phones and fax machines with a new communications infrastructure. It will do nothing further to increase the effectiveness of supply network production scheduling.

The goal of supply network production scheduling is a new system, utilizing the evolving EC infrastructure, that mediates between the parochial PICS of various companies. The result will be a higher level of agility in responding to variances and more efficient control of production across supply networks [see [10] for a detailed review].

In such a system, the role of the human "purchasing agent" will be significantly different. Instead of placing orders, he or she will provide guidance. For example, she may specify the trade-off between long term and short term financial goals, set levels of acceptable risk, and require specific processes for supplier qualification and monitoring. The system will compute production schedules and place orders.

We believe that it is just now becoming possible to implement inter-company PICS that support supply network production scheduling. The purpose of this paper is to lay the groundwork for such a system, tailored to the EC environment, and based on a means of recombining features of prior production scheduling methods.

3. Background

The technical literature on production scheduling falls into two broad categories. One deals primarily with analyses of specific situations (manufacturing states, performance measures, external demands), in order to determine what sequence of operations is expected to lead to the most desirable consequences. The other deals with methodologies and paradigms for production control and their effectiveness in general.

3.1 Specific Situations

In papers dealing with specific situations, the logistical optimization problems occur at various levels of aggregation, ranging from individual machines, to lines, to plants, to enterprises, and even to supply networks. At the line level, a representative problem would be the process optimization of an electronic board assembly line, consisting of several $300K tools, each accommodating up to 100 component feeders; given a set of demands for various board types, the task is to find the assignment of boards to tools and components to feeders, coupled with the sequencing of board types and component placements, that optimizes the overall production [1]. At the supply network level, a representative problem would be the optimum selection of vendors, based on exchange rates and financial and political risks [13].

In addition to dealing with a variety of specific situations, papers report on a variety of approaches. This diversity is partially dues to the complexity of discrete parts manufacturing, which usually makes it infeasible to find optimal solutions even in static situations. The impediment is not a lack of computing capability. Even within a vertically integrated company, the difficulties

include the combinatorial complexity of Bills of Materials, planned change (setup, product ramp-up/ramp-down), unplanned variance (demand, costs, parts, resources, statistical yield), lot size and indivisibility of resources, complex routing (looping), product mix, etc. The objectives are often in conflict, e.g., short vs. long term profit, high throughput vs. small order size, low inventory vs. high order fill rate, high asset utilization vs. high flexibility, etc. Furthermore, the real situations change so rapidly that static problems are at best imperfect formulations, and optimal solutions to static problems are immediately obsolete anyway.

Finding good solutions to these problems may involve the technologies of Operations Research, queuing theory, and Artificial Intelligence. It may also involve imbedding such technologies within interactive decision support tools that provide essential information to human experts who then make the decisions.

3.2 Production Scheduling Paradigms

This paper is concerned primarily with production scheduling methodologies, rather than with the solution to any specific optimization problems.

In response to manufacturing needs, a limited number of production scheduling paradigms have been promoted during the past 40 years. The main ones are Stocking Policy, MRP, and Kanban/JIT [5]. Each paradigm shift has been accompanied by pervasive publicity about its merits, while subsequent experience has uncovered limitations. None of these paradigms is well suited to supply network optimization, especially in the presence of significant demand and supply variance.

The traditional approach is based on a family of diverse "stocking" policies, the simplest of which is Reorder Point [21]. In it, each component part or material has a minimum (Min) and a maximum (Max) level of inventory. When the stock falls below the Min level, enough is ordered to bring it up to the Max level.

Stocking policies have several deficiencies. First, they assume deterministic and/or independent probabilistic demand, ignoring demand dependencies inherent in an exploded Bill of Materials (BOM), such as the correlation of 1 car chassis with 4 tires. Second, when they are used in a multiechelon production process (e.g., a supply chain), an order to restock may incur large delays as it percolates several levels upstream through missing inventory until it triggers the start of production of a needed part. Third, the more sophisticated stocking policies, based on demand forecasts, are mathematically difficult to master effectively. Finally, these policies are sensitive to demand variance. Reorder Point, for instance, leaves dead inventory of at least Min when demand eventually decreases to zero [14][18].

As computers became more prevalent, MRP[1] was conceived to systematically handle the multiechelon BOM problem under probabilistic demand [19]. The approach is particularly suited to assembly. Each product has an assembly BOM, and each item in this BOM has a Production Planning Time (PPT) needed to replace

it. Depending on orders for assembled products, MRP explodes the BOMs backwards in time according to the PPTs, to determine the necessary moment to "release orders to the floor", i.e., schedule the start of production of each component part, down to the lowest level components[2]. The schedules for all parts for all products are then superimposed into a Master Production Schedule (MPS).

Unfortunately, the rationality of MRP unravels, starting with the problem that the superposition is not guaranteed to yield a feasible MPS, since overlapping production activities may conflict over the same resources. To avoid this problem, the PPTs are given a large cushion of safety. For example, if it takes 12 hours raw production time to fill a large order for specific parts, including machine setup, then the PPT might be 30 days. In turn, the long PPT causes MRP to release orders much sooner than actually required. If a PPT were 60 times raw production time, then, very likely, there would be about 60 released orders waiting for that machine. Knowing that the production manager has considerable flexibility in scheduling activity on this resource, customers ask that their particular jobs be "expedited". Expedited jobs then get done, while unexpedited jobs stagnate, causing PPTs to be increased. The longer PPTs cause many more orders to be released, and operations eventually become controlled by an ad hoc system of expediting rather than by MRP. For example, in 1984 one semiconductor plant discovered that 85% of all its manufacturing jobs were flagged as high priority, while most of the rest had been in the pipeline for over a year.

MRP-II was developed to augment MRP with a more objective means of Constrained Resource Planning, but it is subject to the same vagaries of prioritizing. Furthermore, it does not provide a methodology to choose among alternative products when shortages of parts or resources occur. The problem, sometimes referred to as "set management", is that it is better to have products either 100% completed or unstarted than it is to compromise by having all of them 99% completed. The importance of this problem became particularly evident recently when the world's output of semiconductor chips was suddenly curtailed by a fire in a critical epoxy plant; major electronic equipment manufacturers were faced with $100M choices about which products to make and which to delay.

Meanwhile, around the time that many manufacturers were becoming disillusioned with the expensive MRP and MRP-II systems they had installed, Toyota introduced a much simpler system, termed Kanban [17] [24]. In it, as parts flow downstream, cards indicating the need for replacement parts flow upstream. Technically, this system is Reorder Point with Min and Max both equal to 1, implemented with cardboard tags subject to lot sizing. The true novelty of Kanban is its integration into a cultural framework termed Just In Time (JIT), within which systematic reductions in Work In Process (WIP) inventories are used to drive continuous improvement in all aspects of manufacturing. Eventually, WIP inventories come down to a level below which

---

1. Originally termed Materials Requirements Planning, the acronym has come to mean Manufacturing Resource Planning

2. Actually, an "ABC" Pareto classification by cost is often used to distinguish MRP tracked items from common components

variances present unacceptable risks. The implementation of JIT also entails shifting the management of supply inventories onto the suppliers. For example, to meet Toyota's requirements for JIT deliveries of parts, suppliers may find it necessary to have loaded trucks circle the assembly plant until the instant delivery is needed; in Japan this practice has been known to contribute to traffic congestion.

The technical literature generally characterizes MRP as "Push" and JIT as "Pull". Extensive studies and analyses of push and pull have been reported [17] [22] [24]. Some research has shown how system performance may be derived from kanban capacity and production rates [20]. Other work has dealt with the effects of variance in demand and supply [6]. Much of the literature concentrates on lot-sizing, safety stocks, and kanban optimization in multi-stage setups.

The most recent paradigm for production scheduling is based on making Constrained Resource Planning paramount. In particular, the identification of certain classes of bottleneck resource allows mathematical determination of local scheduling and inventory policies to optimize their utilization and off-load their excess work. A proprietary system is the leading example of a decision support system based on this paradigm [11].

The control of inventory is, in fact, the common denominator of all production scheduling paradigms. Inventory is a characteristic of all stages in manufacturing, starting with Raw Materials Inventory (RMI), becoming Work In Process Inventory (WIP), and then Finished Goods Inventory (FGI). As inventory progresses through successive manufacturing stages, its carrying cost increases exponentially (as explained in more detail in a later section). Reducing inventory, or delaying its advance, increases the flexibility of an enterprise by leaving funds available for other uses. Lower inventory, especially at later stages, also reduces losses due to shelf life, shrinkage, and dead inventory when a product is terminated. Conversely, higher inventory provides safety against work stoppages, and later stage inventory enhances the enterprise's ability to respond quickly to increasing product demand.

In summary, Reorder Point captures the trade-off with Min and Max numbers. In MRP, the additional relevant point is that WIP advance is synchronized. JIT also controls WIP advance, but by a needs-based "pull" rather than by a planned "push". In a Bottleneck system, the advance of WIP is encouraged in the vicinity of constrained resources.

3.3 Planned Change vs. Unplanned Variance

Manufacturing can be very simple in a static situation. An example might be the production of buggy whips during the 19th century. At the risk of exaggeration, the product was constant, demand was constant, production was constant, profit was constant, and perhaps this was the last such instance in the entire history of manufacturing. In today's manufacturing world, there are no constants anymore. Instead, there is a continuum of change, varying from gradual, foreseeable changes up to sudden, unplanned changes.

Although most changes are planned, few are gradual. Examples of planned change include the installation of a new line, the replacement of a PICS, or the transition from one product model to another. Although Ford Motor Company planned the changeover from its Model T car to its Model A, the actual changeover took so long that General Motors was able to gain significant market share. A more recent example would be the restructuring of the computer industry to emphasize small computers rather than large mainframes.

In electronic equipment manufacturing, an increasing rate of new product introduction has made management of planned change very important and very complex. But difficult as planned change may be, unplanned variance is much worse.

Unplanned variance wreaks havoc on production schedules. For example, late delivery of parts can cause delay of a production operation, propagating delays to downstream operations. A strike can shut down a succession of plants. Unscheduled downtime of a critical resource has a similar effect. Conversely, early delivery of parts fills up space and compels the manufacturer to perform extra operations of storage and retrieval.

Quality engineering provides methodologies for vastly reducing or eliminating variance due to defective parts. Continuous improvement (e.g., under JIT) is an instance of quality engineering, and a visible result is that modern electronic equipment manufacturing no longer subsumes significant "rework". Nevertheless, in semiconductor manufacturing, statistical yield is an enormous source of variance. During ramp-up of new chips, yields are often below 5%, requiring production demand to be artificially inflated by an uncertain factor of more than 20 to compensate. As a result, a significant fraction of semiconductor manufacturing resources is devoted to products with wildly varying demands.

In electronic equipment manufacturing, demand variance is a pervasive problem. On a long time scale, this variance affects the accuracy of resource plans. On a short time scale, this variance occurs within the lead-time for making (and delivering) a product. Because of strong international competition, manufacturers cannot afford to wait for orders before they commit production. Therefore, they depend on demand forecasts, placing speculative orders and adjusting these orders right up to the moment they take delivery. It is well-known, however, that demand forecasts may be notoriously inaccurate in either direction. For example, the original forecasts for Personal Computers and for a flagship Personal Digital Assistant were both around 50,000 units per year. The first was a factor of 20 too low; the second was a factor of 5 too high.

Independently of the production scheduling paradigm, there is value in accurate prediction of demand. Unfortunately, it is often the case that the forecasting methodology is biased. This bias then skews the results obtained in production planning. For example, Kindergarten through 12th grade demand for computing equipment occurs almost entirely during the summer recess months. A prediction methodology that uses a 4-month floating average will fail to recognize this cyclical pattern. Irrespective of the paradigm for production control, the result is then insufficient WIP in June and excessive WIP in October.

3.4 Supply Networks

The past 20 years have been marked by a pervasive change in manufacturing from vertically integrated manufacturing companies to enterprises that are dependent on manufacturing supply networks. For example, to make a plane with 4 million parts, Boeing Aircraft Company has about 3,000 suppliers, each in turn with dozens to hundreds of suppliers, and so forth.

In Japan, each large manufacturing company is part of a "keiretsu" that includes its immediate suppliers, as well as its financial partners. Because these relationships remain in place for years, the keiretsu provides a framework for concerted information gathering and production control. In the US, some large manufacturers, by virtue of their size, are able to exercise a similar but slightly lower level of power over their suppliers. In recent years, many large manufacturers in the US have begun pruning their lists of suppliers in an attempt to establish long term relationships with a smaller set of qualified suppliers.

Concurrent with the trend towards consolidated control and small numbers of suppliers, the emergence of computer networks is facilitating a trend in the opposite direction. While we believe that the latter trend will eventually win out, no one knows for sure how these conflicting trends will eventually be resolved.

The trends in networking make each company part of a global supply and distribution network. Electronic Commerce (EC) is the latest phase in a process that has been evolving for over a century. Communication among companies has gone through many levels of technology, from mail, to telegraph, telephone, FAX, e-mail, dialed lines, and now Internet. Electronic Data Interchange among companies, pioneered over a decade ago for financial transactions, soon spread to allow exchange of technical data. Applications expanded as supporting technologies for encryption and authentication were developed. One prototype electronic purchasing system has been in operation for over 6 years [8]. Recently the World Wide Web on the Internet has begun to offer a broad range of services, ranging from electronic catalogs of parts [26], [30] to agora style markets [25], [27], [28], [29], [31].

The emergence of EC lowers the barriers to new companies and services, throughout the world. In the EC environment, all of the world's manufacturers and distributors are part of a global network. While individual manufacturers may choose to interact with subsets of this network, the entire supply network is nevertheless available. This transformation is already well underway. Even today for the most part, companies in a supply network are autonomous and competing.

Autonomy and competition imply that access to information and control is limited in a supply network. For any given product, the full state of the manufacturing system is therefore neither knowable nor controllable. This presents a new challenge for production scheduling. To date, production scheduling has been applied primarily to situations in which all essential known production information is accessible, and all control decisions are implementable. The information is collected and analyzed, a control decision is made, and then that control action is taken. This scenario may be applicable within companies that are vertically inte-

grated (or within the military), but it is not applicable to a set of autonomous, competing companies, each of which regards its own data as proprietary and its own control as sovereign[1].

Few papers in the technical literature relate to multiechelon scheduling through supply chains. In auto manufacturing, the impact of EDI has been noted with regard to supply chains [23] as well as quality and inventory [15]. Competition through capacity has also been analyzed [4]. One paper reports on a comprehensive simulation of a large enterprise to study the interdependencies of manufacturing, marketing and R&D [18].

The literature does not address the need for a new paradigm for production control, one that protects proprietary information while exchanging Just-Enough-Information to support cooperative optimization of production across supply networks. The role of EC is to facilitate this exchange.

Although data is exchanged in supply networks today, almost none of the data deals with manufacturing production, e.g., timing of parts and materials, capacities, priorities, demand forecasts, etc. The reason is the huge risk that proprietary information will be used to the disadvantage of any company that provides it to others. For example, a company's confidence in a supplier's ability to meet schedules increases with the information that the supplier has excess capacity; but this same information indicates that the company probably could negotiate a lower cost from that supplier. On the other hand, production data must be shared if the overall supply network is to run efficiently. Without data sharing, variances build up, and parochial production scheduling and control systems can not adequately compensate for these variances.

The current culture of data secrecy will change when a new production scheduling paradigm demonstrates that such change can provide significant economic benefits. At the same time, it will change only if this paradigm provides appropriate new forms of security, probably based on distributed agents, defined protocols, authentication, and cryptography. Successful companies will then use this production scheduling methodology to become more competitive.

The production scheduling policy they use cannot be simply a Stocking Policy (e.g., Reorder Point), because these policies have no knowledge of supply networks.

Similarly, MRP as it is traditionally implemented is not suitable. Because each product's BOM and PPTs are distributed over the supply network, it would first be necessary to collect this data for analysis. If this collection process were done by the final assembly plant, that plant would see information that is clearly not in the interests of all suppliers to provide. For instance, a supplier that simply subcontracts work to a foreign sweatshop may not want to reveal a BOM with only one branch.

To the extent that Kanban is equivalent to Reorder Point, it also is inappropriate for supply networks. Additionally, since the essence of Kanban is the use of tags to

---

1. Even in large integrated companies, different divisions or subsidiaries all too often behave as if they were autonomous and competing.

avoid computer tracking, the necessary production data would be unlikely to be on-line. JIT is not relevant, since it philosophically transfers problems of WIP management to upstream suppliers, rather than dealing with the global optimization problem.

Bottleneck paradigms are inappropriate, since owners of bottleneck resources may not want to reveal the huge opportunity for competitors. Conversely, underutilized suppliers may be reluctant to provide capacity data because it would undermine their negotiating position in future contracts (and because it could lower their price on the stock market).

4. New Just-Enough-Information Paradigm

This paper proposes a specific way of combining elements from prior policies into a new JEI production scheduling paradigm that is more applicable to manufacturing supply networks. It is based on the observation that in a supply network, the benefits of MRP can be achieved by a recursive procedure that we are calling Demand-Availability-Order (DAO), which does not require any supplier to provide proprietary data. DAO is an algorithm that can support a variety of production scheduling policies. These policies, in general, have the benefits of prior systems while avoiding the disbenefits.

DAO and a scheduling policy are only part of the new JEI paradigm. This paper also discusses other parts that we believe are necessary, including pricing, qualification, negotiation, and capacity management. There are also some related security issues. Much further work is needed on these other components before the entire new paradigm for supply network production scheduling can be articulated. The primary purpose of this paper is to explain the operation of DAO, offer some evidence that it works, and introduce the other topics.

4.1 Demand Availability Ordering

We begin with three simplifying assumptions: single product, single source, single customer. In later sections we discuss how these assumptions can be transcended.

Consider a situation in which a manufacturer is already making a specific product to meet the downstream demand of one or more customers. The manufacturer makes this product by assembling components provided by a set of immediate upstream suppliers. Assume that the product is in mid-life, i.e., the overall supply network has capacity allocated to this product and there is already inventory (RMI, WIP, and FGI) in appropriate places.

We define a "schedule item" to be an ordered pair {quantity, date} consisting of a quantity of a given item and a date, which may be past or future. A "schedule" is a list of schedule items for a specific product. A "demand schedule" is a schedule that is sent by a manufacturer to an immediate upstream supplier, expressing the manufacturer's past and anticipated future need for that item over time. An "availability schedule" is a future schedule, consistent with a "current" demand schedule, that expresses a supplier's ability to deliver

that item over time. An "order schedule" is a future schedule, consistent with a current availability schedule, that expresses a manufacturer's commitment (i.e., unilateral restriction) to accept delivery from an immediate upstream supplier. An order commitment completes the "binding agreement" by assuring that the supplier intends to make that delivery. "Current" and "binding agreement" are concepts that we discuss more fully in a later section.

In the DAO algorithm, a product manufacturer is the initiating node in a supply network. This node sends demand schedules to all immediate suppliers of components for that product. It gets back their availability schedules. It then combines these availability schedules into a "worst-case" availability schedule for these components, i.e., a schedule that assures complete component sets for its product. It converts this into order schedules and sends them to the immediate suppliers. They send back order commitments.

The concept of "worst-case" availability is based on the assumptions that a supplier can always shift availability to a later time, and that lot sizing is not a problem. We examine these assumptions more fully in a later section.

The DAO algorithm proceeds recursively at each node of the supply network:·

- When a node receives a demand schedule then if it is a terminal node, it returns an availability schedule. Otherwise, it sends derived demand schedules upstream to its immediate suppliers and waits for them to return availability schedules.

- When a node receives all upstream availability schedules, it combines them to obtain a single worst-case availability schedule. If this node is not the initiating node, then it adds on its own PPT to derive its own availability schedule and sends it downstream.

- When a node receives an order schedule, then if it is a terminal node, it returns a commitment. Otherwise, it sends derived order schedules upstream to all immediate suppliers and waits for them to return commitments.

- When a node receives all upstream commitments then if it is not the initiating node, then it sends a commitment downstream.

For example, suppose that the supply network consists of one manufacturer M with 2 immediate suppliers W and X, and X has suppliers Y and Z. Then:

1. M sends demand schedule D_w to W and D_x to X.

2. W and X then act in parallel:
   a) W sends availability schedule A_w to M.
   while:
   a) X derives demand schedules D_y and D_z and sends them to Y and Z.
   b) In parallel: Y and Z send availability schedules A_y and A_z to X.
   c) X combines A_y and A_z into worst-case availability schedule A_{yz}.

d) X derives availability schedule A_x and sends it to M.

3. M combines A_w and A_x into worst-case availability schedule A_{wx}.

4. M derives order schedules O_w and O_x and sends them to W and X.

5. W and X then act in parallel:

   a) W sends commitment to M.
   while:

   a) X derives order schedules O_y and O_z and sends them to Y and Z.

   b) In parallel: Y and Z send commitments to X.

   c) X sends commitment to M.

4.2 Features

The salient feature of DAO is MRP equivalence, i.e., when a new order is placed by DAO, all releases to the shop floor occur at precisely the same time that they would under MRP.

Unlike MRP, however, DAO hides each node's PPT and BOM from all other nodes. (Of course, a manufacturer is free to share parts of the BOM with suppliers on a need-to-know basis.) Also, DAO in a supply network is much less susceptible to PPT inflation than is MRP in its typical application. The reason, which to some degree has little to do with DAO vs. MRP, is that in a supply network each supplier has a competitive incentive to keep its PPT as low as possible. It knows that the customer can always switch to a competitive supplier, and it therefore does not want to provoke a bidding war based on earliest availability.

Another feature is that DAO is consistent with accountability for supply variances. In particular, DAO specifies committed time to deliver, while MRP specifies committed time to start, i.e., release orders to the plant floor. In a later section, we will discuss how DAO accountability could be formalized in a downstream alert system. Additionally, while MRP only deals with orders, DAO provides a natural means to deal with forecasts as well.

DAO is compatible with actions taken by individual suppliers to provide local safety stocks, over the levels explicit in their customers' orders. Furthermore, DAO could be extended to include a concept of "speculative orders" that could provide an interesting new formal means of risk management with regard to increases in demand. This concept also is discussed in a later section.

4.3 Implementation

We anticipate several stages to the pervasive introduction of any new paradigm for supply network production scheduling. The first phase is to write software for an abstract model of a supply network, use it to demonstrate general characteristics of the policy, and collect evidence that the policy outperforms prior methodologies. The second stage is to design and build the control structures for a real system (generic database schema, database access calls, protocol for communication, communication calls, etc.) and simulate their operation in a single processor multitasking system. The third phase involves porting the system to a set of networked processors, representing the PICS of a set of autonomous companies. The fourth phase is to apply the system to a real but small manufacturing supply network. Ultimately, we anticipate that this software would be imbedded into a variety of current and future PICS.

To date, we have been working only on the first stage. Our results are described in the next section.

5. Experimental Results

A supply network may be represented as a directed graph, with factories as nodes and parts transportation as arcs. Some of the nodes represent sources of raw materials, and a sink node represents the final customer. Because we are considering only assembly and we are assuming no rework, the graph is acyclic, and the subgraph of all the assembly factory nodes is a tree. Each assembly factory has a set of one or more inputs and one output.

Within each factory, the inputs and the output are all queues with maximum and current lengths. Each factory has an associated assembly time. Similarly, each part transportation arc has an associated transportation time. (The sum of the assembly time and the output transportation time represents that factory's PPT.)

To complete this model, some representation is needed to deal with the extent to which each factory is capable of multiprocessing, i.e., able to be in different stages of peristaltically processing several assembly jobs that overlap in time.

5.1 Abstract Model of Supply Network

In order to obtain results rapidly, we use the simplest abstract model that we feel adequately represents the essential aspects of the supply network production scheduling problem. Our simplifying assumptions are:·

- The factory tree is balanced, i.e., every factory node has the same number i of input arcs, and each branch of the tree has the same number of levels l of factory nodes.

- The queues all have equal capacities, i.e., every input and every output queue has a maximum length b.

- The assembly time for each factory plus the transportation time for each arc add up to one time unit. Furthermore, all factories are in lock-step synchronization, i.e., during each unit time cycle, all factories do a "make" operation and then all transportation arcs do a "move" operation. This sequencing deliberately precludes the possibility of work unrealistically moving through more than one make operation during a given cycle.

Earlier, we defined a schedule as a list of pairs {quantity, date}. In the abstract model, however, all dates can be replaced by small integers that represent the cycle number. Furthermore, there is never a need to

deal with a past cycle older than the oldest unfilled order or a cycle more than *l*+1 steps into the future. As a result, the representation {quantity1, quantity2, ..., quantity n} is an equivalent but simpler representation for a schedule, provided it is clear from the context how to map the indices onto the cycle numbers.

Parts supply variance is represented by having the sources for raw materials generate outputs according to independent probability distributions that are uniform from 0 to *b*. (We plan to replace this model eventually with a more traditional one in which each factory may be up or down during each cycle, as determined by independent probability distributions.)

Customer demand is represented as a sine curve, with period *t* = 12 time units (e.g., months) and mean value λ' oscillating between 0 and a peak amplitude of 2λ'. The value λ' is adaptively tuned to actual mean throughput λ (over a period much longer than 12 time units) by increasing λ' by a factor of 1+ε whenever the current order can be filled from the final factory's output buffer and by decreasing it by 1-ε whenever it cannot. We use ε = 0.1, after verifying that the model is insensitive to the exact value of ε. (This damped response of the mean implies that the model also represents a situation of exponentially increasing or decreasing demand with a characteristic growth/decay time constant much longer than 1/ε time periods.)

The model is typically used in a Monte Carlo simulation [see for example [12]]. Each time cycle, the simulation proceeds through the following steps:

1. Customer: The end customer imposes a demand.

2. Schedule: The production scheduling policy is applied, and the resulting production orders are received at each node and arc.

3. Execute: The production orders are executed:

   a) Make: Factories do assembly, moving inventory from inputs to output queues.

   b) Move: Inventory is transported from output to input queues. (At the downstream end, FGI is moved to the customer; at the upstream ends, RMI is acquired.

4. Monitor: The state of the overall system is examined to determine incremental performance.

To support the above steps, the model also provides a means of representing pending production orders at each node and arc.

The behavior of this simulation is determined entirely by the production scheduling policy and only four free parameters (*l*, *i*, *b*, and *t*). The small number of parameters is an asset, since, in accordance with Occam's Razor, it precludes the possibility of our deliberately or inadvertently tuning the system to bias the results. We explore the parameter space:

levels: *l* = 1, 2, 3, 4, 5
queue size: *b* = 1, 3, 10, 30
inputs: *i* = 1, 2, 3, 4, 10
period: *t* = 12

For each set of parameters, the simulation is started and run until transient behavior has damped out. Then it is run for a large number of trials (~10,000), during which data is collected. At the end of the run, the data is analyzed. Finally, the data associated with different production schedulers is compared.

In performing the analysis, it is essential to have some measures for *C*="inventory carrying cost" and *P*="value of product shipped". Recognizing that real world companies may have strong and differing views of how these terms should be measured, we find it necessary nevertheless to impose what may be perceived as ad hoc choices:·

- We assume that each raw material has an input cost of $1 per item.

- We assume that the Value-Add for assembly operations is equal to the Cost of Materials. (According to 1991 US Department of Commerce numbers, for the combined manufacturing sectors of Computing and Office, Household Audio/Video, and Communications Equipment, Value-Add was about 96% of the Cost of Materials, so this is a very good approximation.)

- We assume that the Value-Add for transportation operations is equal to 1/2 of the cost of materials transported. (We have no numbers to support this, and we suspect it is an overestimate, but it is a convenient number for purposes of computation.)

Based on these assumptions, it is straightforward to compute the "value" of any inventory item, whether it is RMI, WIP, or FGI. For example, if *i* = 3, then 3 raw material items of $1 each are assembled into an output item with a value of $6. At the next level downstream, the 3 inputs each have a value of $9, and the output item has a value of $54. It can be seen that the value of an inventory item rises exponentially at each downstream level. At each instant, we define inventory carrying cost *C* as the sum of the values of all inventories, computed in this manner. Similarly, value of product shipped *P* is the value of FGI, computed this way. (The actual expense associated with inventory would be proportional to *C*, based on amortization; the proportionality constant would represent the prevailing discount rate plus an allowance for inventory shrinkage and dead inventory.)

It is also useful to have a measure of "total WIP inventory", since this appears as the term *L* in Little's Law $L = \lambda W$, where λ is throughput and *W* is average throughput time. However, it is not correct to let *L* be simply the sum of all RMI, WIP, and FGI, with each item counting as 1. Instead, *L* is a weighted sum, with each item weighted according to the fraction of FGI that it ultimately can contribute. Thus, each item of FGI carries weight 1. At the next upstream level, each item of input (or output) WIP carries a weight 1/*i*. At the next level the weight is $1/i^2$, etc.

5.2 Policies Evaluated

Within the abstract model, any production scheduling policy can be succinctly characterized. For the policies we chose to test, the following variables are relevant:

D = demand schedule = { $d_1, d_2, ..., d_l$ }
A = availability schedule = { $a_1, a_2, ..., a_l$ }
O = order schedule

We tested four policies, the first two of which are not compatible with DAO because they order the advance of WIP independent of the availability of correlated parts:

1. "push": push materials everywhere at the maximum possible throughput. Start as much as possible; order all WIP to advance if possible.

2. "pull": pull materials everywhere at the maximum possible throughput. Start just enough more to fill the output queue; order all WIP to advance if possible.

3. "mwa" (Minimum WIP Advance): Start just enough more to fill the output queue; order WIP to advance only when necessary. D = { $b, 0, ..., 0$ }; O = { $a_1, a_2, ..., a_l, b - (a_1 + a_2 + ... + a_l)$ }.

4. "mwas" (Minimum WIP Advance with safety stocks): The suffix "s" in the acronym indicates a safety stock policy, obtained by modifying policy 3 to have all terminal factories provide a safety stock {b-a} of all input raw materials.

In addition to the above policies, we also simulated "dmwa" (Demand-based Minimum WIP Advance), in which the first element of D is total actual demand d instead of b, and the last element of O is similarly modified. This policy starts just enough more to satisfy known demand. We also tested "pmwa" (Prediction-based Minimum WIP Advance), which starts just enough more to satisfy known demand d plus predicted future demand p. Additionally, we tested "dmwas" and "pmwas", the safety stock variants of dmwa and pmwa. The behavior of mwa, dmwa, and pmwa was found to be highly similar, as was the behavior of mwas, dmwas, and pmwas. For this reason, we are minimizing further discussion of these other 4 policies.

5.3 Enumeration Results

For the (l,b,i) parameter sets (1,1,2) and (2,1,2) the number of possible states of the overall supply network is relatively small. The number may be further reduced by grouping states that are equivalent under certain symmetry rules.

We used these principles to write short computer programs that enumerate all (1,2,2) and (2,1,2) states and all transitions in order to obtain exact results for the "push" and "pull" policies. The main value of doing this was that it provided an independent means to check the simulation results for the same parameter sets. The result is a higher degree of confidence that the simulation is correct for all policies and all parameter sets.

5.4 Analytical Results

For the (l,b,i) parameter sets of the form $(l,1,i)$, we have also written symbolic expressions, in the form of infinite series, that provide the exact λ under mwa for a slightly different model. Unfortunately, we have not yet been able to find closed form solutions except in a few limiting cases. Nevertheless, we have written a program that quickly computes numerical answers. The quantitative and qualitative agreement of these results with those of the simulation further adds to our confidence that the simulation is correct. Additionally, this work allows us to get numerical results for large i and large l, results that are infeasible to obtain in a reasonable time by simulation.

5.5 Simulation Results

To perform the Monte Carlo simulation, we considered writing code either in an existing simulation language (e.g., SIMAN, SIMSCRIPT-II.5, SLAM, ithink)[1], or in a conventional programming language (e.g., C or C++), or in a combination of the two. We recognized that the facilities of simulation languages make them far superior for modeling specific complex networks. On the other hand, for the relatively simple abstract model above, they offered no obvious advantage over C++, and they were costlier and slower. Additionally, we were concerned that in the long term, as we progress towards implementation of systems in real factories, these more powerful simulation systems might prove considerably harder from the point of view of networked communications, interfaces to a variety of users' environments, and access to users' legacy information systems. For these reasons, our work is in C++. (Actually, as we come up the C++ learning curve, we have been working in a somewhat unconventional C-based environment on an interim basis.)

In running simulations, we measured throughput λ, total WIP inventory L, carrying cost of inventory C, value of product shipped P, and mean delay w to order fill. From λ and L, Little's Law permits us to infer mean throughput time $W = L/\lambda$.

Our results indicate no qualitative differences in behavior for i from 2 to 10, as well as no qualitative differences in behavior between b = 1, 3, 10, and 30. The results also indicate that the variation as a function of l is continuous and not rapidly changing. For $b \gg 1$, we observe that for each paradigm, λ is linearly proportional to b, although the constant of proportionality may depend on l and on the paradigm. This proportionality is expected, since the model assumes that each factory's capacity is limited by the capacities of the input queues. (For small b the proportionality tails off because random fluctuations are more likely to leave an empty input queue.)

For i=1, the results of pull and mwa are consistent with $\lambda = b/(l+2)$. The denominator $l+2$ in the proportionality constant is the average number of time cycles between the time raw material is requested at the upstream end and finished product emerges from the downstream end. $l+1$ of these cycles are due to l make and move operations plus 1 raw material delivery cycle.

---

1. SIMAN is a product of Systems Modeling Corp., Sewickley, Pennsylvania; SIMSCRIPT-II.5® is a product of CACI Products Company, La Jolla, California; SLAM is a product of Pritsker Corp., Indianapolis, Indiana; ithink™ is a product of High Performance Systems Inc., Hanover, New Hampshire.
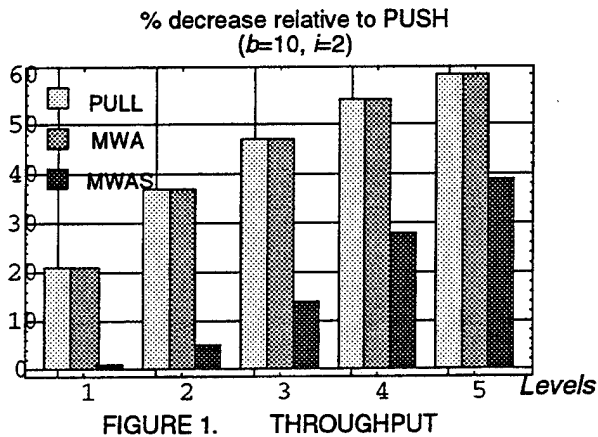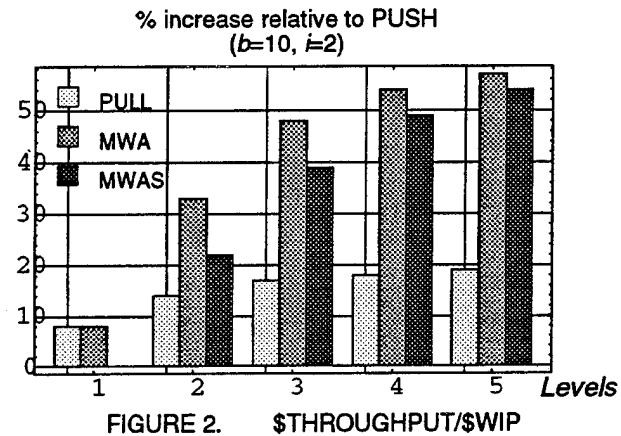
FIGURE 1. THROUGHPUT

FIGURE 2. $THROUGHPUT/$WIP

The remaining 1 cycle is the average time delay in getting raw material. (This delay is due to the model's assumption that raw material input has a uniform distribution. On the first cycle, a demand for b yields an average input of only $b/2$; on the next cycle, a demand for $b/2$ yields $b/4$, etc.)

For $i = 1$, the results of push is consistent with $\lambda = b/2$, i.e., a proportionality constant of 1/2, independent of the number of levels l. This behavior is attributable to the model's assumption that raw material input has a uniform distribution from 0 to b. For push, this average raw material input of $b/2$ in each cycle simply advances peristaltically, so an average of $b/2$ comes out every cycle. For mwas the results are consistent with $\lambda = b/2$ for $l=1$, transitioning to $\lambda = b/l$ for $l >> 1$.

Because we see no qualitative differences in behavior as a function of $b >> 1$, $i > 1$, and l, we feel free to present representative simulation results, rather than all our results. Ideally, the results we choose to present should be based on (l,b,i) parameters that are characteristic of real supply networks. Unfortunately, a search for relevant survey data (e.g., in [1]) has failed to find information on the structure of supply networks in real commerce. We are, therefore, creating a survey on Internet (URL: http://irobot.isi.edu/flex/flexmain.html) to collect such data. We expect our survey to determine eventually that in real supply networks i is on the order of 10, b is on the order of 1000, and l is roughly 4.

In the meantime, we choose to present simulation results that focus on $i = 2$, $b = 10$, and $l = 3$. This set provides a good expository example, and it is likely to be extremely conservative in assessing the ultimate value of DAO in real supply networks.

Figures 1 and 2 compare the performance of pull, mwa, and mwas to push. (The statistical uncertainty in these results is around 1-2%.) Figure 1 shows, as expected, that all other policies result in lower λ than push. Figure 1 shows, as expected, that all other policies result in higher P/C than push.

Consider the 3 level supply network results. The numbers for throughput λ and cost ratio P/C for pull, mwa, and mwas relative to push are:

- pull: a decrease of 47% in λ; a gain of 17% in P/C
- mwa: a decrease of 47% in λ; a gain of 48% in P/C

- mwas: a decrease of 14% in λ; a gain of 39% in P/C

The numbers for mwas are quite good, i.e., not much degradation in throughput and considerable improvement in cost ratio. For the same set of parameters, the other results comparing mwas to push are:

- a decrease of 17% in total WIP inventory L
- a decrease of 4% in average throughput time W
- a decrease of 12% in mean delay to order fill w

From our simulation results in general, we draw the following conclusions:·

- Push achieves the highest throughput. (This is not surprising, of course.)
- Minimum WIP Advance policies without safety stocking (mwa, dmwa, pmwa) perform roughly equally.
- They are considerably worse than push in λ, but considerably better in P/C.
- Minimum WIP Advance policies with safety stocking (mwas, dmwas, pmwas) perform roughly equally. They achieve nearly the same λ as push and nearly the same P/C as their counterparts without safety stocking.

Overall, the conclusion is that DAO works, that the performance of policies with safety stocking implemented under DAO are remarkably good, and that this behavior holds for supply networks of all levels. These results give confidence that the method should be pursued further.

6. Generalizations for Practical Application of JEI

6.1 Pricing and Qualification

The field of game theory deals with interactions of separate players, each of whom is optimizing his own interests under the rules of the game [see for example

[9]]. By anticipating the likely behavior of the players, one can set up the rules of the game in such a way that they incent or disincent certain patterns of behavior. When game theory is applied to economics, the relevant subfield deals with equilibria in multi-period non-cooperative games.

This situation applies to autonomous companies in commerce. For conventional commerce, the rules of the game have been evolving for millennia, and they continue to evolve. For Electronic Commerce, the rules have barely begun to be articulated.

For production scheduling of manufacturing supply networks under JEI, this issue surfaces as soon as generalizations are considered. Therefore, we must define rules of the game such that the selfish interests of the separate autonomous companies do not destroy the benefits of cooperative behavior. The relevant rules of the game are pricing and qualification, both of which are well-known in conventional commerce, and both of which are amenable to technical solutions under EC. Regulation is another mechanism, but since it is generally non-technical, it is outside the scope of this paper.

Pricing would associate certain payments with DAO steps, based either on global agreement or on bilateral negotiated agreements. Pricing alone is insufficient, however, because a company may cause damages to another company that far exceed the cost of any part[1]. We believe that the solution to this problem will take the form of qualification services, provided by independent third party organizations (analogous to credit reference companies). These service companies, which are already coming into existence under EC for purposes unrelated to production scheduling, will be authorized by suppliers to monitor their DAO activity and their actual performance, prepare reports, and make the reports available to other authorized manufacturers.

In the discussion of DAO generalizations below, we will indicate the problems that arise and indicate how the mechanisms of pricing and qualification might provide solutions.

6.2 Lot Sizing and Delayed Availability

In an earlier section, we noted that the synthesis of availability schedules into a single worst-case availability schedule implicitly assumed that availability can always be shifted to a later time and lot sizing is not a problem. A specific example illustrates these issues.

Suppose that an assembly consists of two parts A and B from two suppliers, and suppose that their respective availability schedules are:

- Part A: 100 at time 1; 110 at time 2
- Part B: 110 at time 1; 100 at time 3

When these are combined into a single worst-case availability schedule, the result is:

- Asm A+B: 100 at time 1; 10 at time 2; 100 at time 3

This schedule, however, is based on the assumptions that 10 units of part A can be delayed from time 2 to time 3, and 10 units of part B can be delayed from time 1 to time 2. Also, it is based on the assumption that

---

1. "For want of a nail the shoe was lost, for want of a shoe the horse... the rider... the battle... the kingdom."

10 is a feasible lot size, even though all the lot sizes in the original availability schedules exceeded 100.

In real manufacturing, the supplier's minimum lot size may be determined by technology, economics, or a combination of the two. For instance, in semiconductor manufacturing there may be several hundred "die" on each wafer, and technology does not allow processing of anything smaller than a wafer. In card assembly, a lengthy machine setup may preclude setting up a component insertion machine for a very short production run. Lot sizing may also be an issue for the customer. For instance, the customer for components may insist that they be provided in rolls of 1,000.

For these reasons, it may be necessary to augment schedules with a specification of minimum lot size. The recipient of the schedule is then able to respond in a consistent manner. In the example above, if minimum lot size were 50, then the worst-case availability would have to be further delayed to give:

- Asm A+B: 100 at time 1; 110 at time 3

In real manufacturing, however, there are also problems with delayed availability, since it can be achieved only by holding WIP or by postponing production. Depending on whether the supplier or the customer holds the WIP, one or the other would incur the associated WIP carrying cost. Regardless of who pays this cost, the result would probably be to raise the downstream price of subsequent assemblies. The mechanism of pricing, as mentioned above, can be used to disincent delays.

Additionally, there are two related issues of capacity management. One is that holding WIP requires storage space and labor, both of which are resources. The other is that production can be postponed only if the plant's capacity can accommodate the shifted production. Capacity management is discussed in a later section.

6.3 Change and Variance

Change and variance refer to the fragility of the supposedly "binding agreement" that DAO provides between a supplier and a customer. Planned change affects DAO if the change is planned to occur within the time horizon of any supplier's longest commitment. A relevant example would be the decision by a supplier, based on gradual deterioration of a critical piece of equipment, to perform previously unplanned maintenance at some future date. Although this decision may be made long in advance, previous commitments may be impacted by lack of future availability of that resource. In making the decision, the supplier is able to compute the extent of change in its committed deliveries. The effect, if any, may be to make some deliveries early, some late, and some infinitely late.

Unplanned variance might be caused by lack of maintenance above, if the critical resource suddenly fails. The supplier then computes the effect on committed deliveries. Alternatively, unplanned variance might be a sudden change in customer demand.

6.3.1 Supply Variance

Supply variance can be dealt with both proactively

and reactively.

A manufacturer anticipating potential supply variance can proactively provide a safety stock of parts (and a safety capability for manufacturing). Subsequently, if a variance occurs, the safety stock can be used to absorb the impact, so that downstream customers are shielded from it. Safety stocks, however, represent WIP, which has significant carrying cost. Each manufacturer, therefore, faces an inventory management problem that trades off between safety and minimum WIP.

A plausible ad hoc method would be for a manufacturer to set safety stock levels of input parts at a percentage of their known demand as a function of time. A virtue of this approach is that in principle it will leave no dead inventory when a product is eventually phased out. This approach (or any of the mathematical techniques described in the literature on Stocking Policy inventory management) is implementable under DAO by simply padding upstream orders.

The reactive approach to supply variance deals with "alert" signals. Whether the cause is planned or unplanned, when a supplier knows that its order commitments have become infeasible, it has an obligation immediately to alert to its immediate downstream customer. In general, a supply variance alert is simply a revised availability schedule. A factory receiving such an alert determines the impact on its own committed orders. If there is an impact, then it passes the alert to its customer along with a revised worst-case availability schedule, maximally compatible with the prior committed order schedule.

Among all alerted factories, the one that is most downstream then responds to the alert availability schedule either with a modified order or by initiating a new DAO process. One or the other is necessary because it reestablishes correlation of WIP advance.

The absence of a supply variance alert does not imply the absence of a supply variance. As an extreme example, a plant that burns down is unable to deliver a downstream alert. Customers, therefore, should periodically initiate upstream order "verifications" to be sure that asynchronous events have not been lost.

Game rules can be used to incent desirable behavior with respect to supply variance. There are two relevant cases: If the primary cause of the variance is a truly unpredictable event (e.g., the Kobe earthquake), then pricing should be designed to reward the existence of prior safety stocks that reduce solve the variance, rapid propagation of significant alarms, and filtering of minor alarms to avoid "churning". On the other hand, if the variance were caused by foreseeable inaccuracy in an earlier availability schedule, then pricing should penalize the source of this inaccuracy. Additionally, qualification services should detect systematic patterns of availability inaccuracy, resulting in penalties for the responsible company.

6.3.2 Demand Variance

Order padding, mentioned above, is also a proactive way of managing risk with regard to sudden increases in demand, although it aggravates the inverse problem of demand decreases.

Reactively, the response to any change in customer demand is the initiation of a new DAO process. As the new demand schedule proceeds recursively upstream, each supplier must be provided with sufficient information to be able to recognize that this new demand is associated with an old committed order. This association allows the supplier to take immediate speculative action, based on its estimation of the likelihood that the new demand is a precursor to an actual order change. Also, the supplier is able to factor out any capacity load due to the old order when it is constructing a new availability schedule.

A factory receiving such a demand change determines the impact on its own committed orders. If there is an impact, then it passes the demand upstream. Otherwise, it absorbs the change. Game rules can be used to discourage order changes, especially changes at the last minute.

6.3.3 Multiple Products, Sources, and Customers

Up to this point, the discussion of DAO is based on the assumption that at every factory there is only a single customer, a single output product, and a single source for each component. In the real world, however, there would be multiple products, sources, and customers. Allowing these generalizations introduces many complications, which are discussed in the following sections.

6.3.4 Asynchronous Transactions and "Current" Schedules

With multiple products, sources, and/or customers, suppliers would receive and send a large number of concurrent asynchronized unrelated demand, availability, and order schedules and commitments. With all of this activity occurring, the state of the overall manufacturing system changes in real time, so that old data may no longer be relevant.

Following receipt of a demand or order schedule, the time taken by a supplier to derive and send subordinate schedules is likely to be short. On the other hand, the latency time before receiving back all availability schedules or all commitments would be much longer and indeterminate. It is likely that the state will change during the time.

We are all familiar with this phenomenon in the context of airline reservations. If one phones for information and later calls back to book arrangements, there is a likelihood that the information is no longer "current" because hundreds of other people were meanwhile transacting state changes. To address this problem, airlines typically hold provisional reservations for a limited period, hold guaranteed or paid reservations indefinitely with a penalty for cancellation, and overbook.

Demands that are concurrent and asynchronized force a consideration of capacity, reliability, and risk. For a typical flight, the probability of any one booking being canceled is around 1/2 and the number of independent bookings is around 100, leading to statistical fluctuations of roughly 10%, which is acceptable. On the other hand, if a single travel agent wished to book 75% of all available seats on a flight, then, even with evidence of the agent's reliability, the airline might be concerned about statistical fluctuations. Even though

that single booking might represent very high profit, the risk might warrant a separate chartered flight.

Of course, the analogy between airlines and manufacturing is imperfect. In manufacturing, typical terms are net payment 30 days after delivery, whereas airline tickets are usually paid before the service is provided. Also, manufacturing commerce has no institution analogous to the FAA that enforces strict equality of all customers or that defines procedures for mediating when capacity is overbooked. Nevertheless, similar techniques apply in these two domains.

Part of the solution under DAO involves time-stamping each schedule when it is sent.

This time is associated with a concept of "validity". At time 0, the validity is 1; for later times it is a monotone decreasing function which is asymptotically 0. For example, the function might be valid for 5 days and invalid afterward. As long as related transactions occur within time periods of high validity, the presumption is that all commitments are firm.

The concept of validity would also encourage suppliers to provide "unsolicited" availability schedules. The analogy is with current business practice of providing, for example, an Autumn catalog, with no guarantee that the same items will be available in Winter.

Pricing is another part of the solution. Pricing can be used to give suppliers an incentive to respond quickly, and to monitor and control their capacity in an intelligent manner. (Capacity management is discussed in a later section.) When submitting a demand or order schedule, the customer makes a payment consisting of two parts: a fixed non-refundable fee, and a fee for fast response. Depending on the time taken by the supplier to respond (with availability or commitment) the supplier retains only the validity fraction of the fast response fee and refunds the rest. The supplier further refunds a portion of the retained fee depending on how different (according to some metric for similarity of two schedules) availability is from demand, or if commitment is not possible on an order.

6.3.5 Variance

Multiple customers and multiple suppliers introduce problematic new dimensions to variance. There are two situations, termed divergent and convergent, that are not necessarily mutually exclusive:

In the divergent situation, a given part may be used by a supplier in a product (or more than one product) for more than one customer. Suppose the supplier receives an alert about impending delayed delivery of such a part. Then the supplier needs a methodology to decide which customer to satisfy first. Faced with such a situation, a supplier will tend to use an informal priority system that maximizes self benefit. A big customer is perceived as more important than a small one, a high visibility customer more important than a low visibility customer, etc.

If we assume that all information needed to make this decision should be available locally, then we are driven to invent a formal prioritizing system in which customers have some opportunity in advance to paid a fee for priority. Prioritizing of customers for this purpose, however, may be viewed as a specialized instance of capacity allocation, in which the resource being allocated is "shipping-dock" capacity. Capacity allocation is discussed in a later section.

Unfortunately, the information needed to choose which customer to satisfy may not be adequately captured by a local priority system. Even though the products may be destined for different customers, it may be that further downstream all of these products come together as components in a single product. As mentioned earlier, set management is very important downstream in final assembly, and one of the main benefits of DAO is that it is supposed to exploit such parts correlation to avoid premature WIP advance. Therefore, if the products ultimately converge, the manufacturer responsible at the point of convergence must be involved in the allocation. This observation suggests the need for header information in the order (and demand) schedules that provides an "audit-trail" of the chain of downstream customers in some encrypted form, so that a supplier knows of the existence of downstream convergence even though it does not know the identity of that manufacturer.

In the convergent situation, a given part may be obtainable by a manufacturer from more than one supplier. The manufacturer had enormous freedom in issuing demands and orders for that part. If it is now alerted to a supply variance by one of its suppliers, the manufacturer has even greater freedom. It can try to shift the order shortfall to other suppliers, or it can shift all orders anew, or it can pass on the shortage to its customer(s), or a combination of the above.

Compounding the difficulty of this choice is the possibility that the variance is caused by some supplier further upstream, and that the variance from this cause is propagating down different paths at different speeds. Thus, hasty action by the manufacturer to shift orders may actually penalize the supplier who most rapidly passed the alert downstream, i.e., the most responsible supplier. This observation suggests the need for header information in the alerts and availability schedules that provides the chain of upstream suppliers in some encrypted form, so that a manufacturer knows of the existence of upstream divergence even though it does not know the identity of that supplier.

Both the divergent and convergent situations, therefore, would benefit from an encrypted audit-trail header. Such a header could also be used to provide a mechanism that expedites non-filterable "direct" communication between non-adjacent suppliers. In the first situation, for example, this feature would allow the divergent shortfall supplier to target a non-filterable alert directly to the convergent downstream manufacturer. To preserve anonymity, the alert would pass physically via the topologically intervening manufacturers. The originator would know only that its alert was directed to, say, the 4th customer in that header chain. The final recipient would know conversely that the alert came from the 4th supplier in that chain. Of course, there is great commercial risk in allowing arbitrary messages to flow in this manner, and we have not fully thought out these issues.

6.3.6 Shopping Around

In selecting a supplier, the most important considerations are price, delivery, and qualification. Price would

be quoted through conventional means. DAO offers a simple mechanism to learn the delivery times of several competing suppliers. A manufacturer would send the same demand to several suppliers with the intent of choosing only one, based on its price and availability "bid" as compared to all others. (There might also be room for some enhancement to DAO to facilitate availability negotiation rather than bidding.) If all suppliers provide honest availability responses to all demands, a "perfect" market results. This market would self-regulate as the demand for rapid availability would either drive up its price or consume its capacity, until non-cooperative equilibrium was reached.

This system has the same sorts of issues that occur in conventional commerce. Suppliers would know that a demand is not always followed by an order, so they might elect not to respond to all demands. In particular, a supplier would probably prefer not to respond to a demand from a competitor. However, given the proliferation of Internet sites[1] it may be quite hard to determine whether or not the demander is a competitor. (A popular cartoon shows a dog typing on a computer keyboard while explaining to another dog that "on Internet, no one knows you're a dog".)

As mentioned earlier, qualification services would disincent suppliers from willfully providing dishonest availability schedules. Pricing would not suffice for this, since any penalty for failure to meet commitments would usually be too little too late. From the point of view of a final assembler waiting for a $1 part to complete a $1M computer, even bankruptcy may be an insufficient punishment for the bad supplier.

Similarly, third party qualification services would provide suppliers with information on the frequency with which a potential customer submits demands with no intent to order. Potential suppliers will utilize these services when deciding whether or not to respond to specific demands. This approach would also defeat attempts of one company wishing to gain information about its competitors through the simple subterfuge of sending them frequent demands.

6.4 Security

The administration of DAO raises some security problems within the supply network. As mentioned earlier, there is probably a need to include an audit-trail header in each demand, availability, and order schedule and commitment. The need is for a supplier to know of the existence of a convergent downstream manufacturer, and vice versa for a customer to know of the existence of a divergent upstream supplier. But, elements of this header require encrypting to prevent suppliers from knowing the identities and needs of non-immediate customers, and to prevent customers from knowing the identities and capabilities of non-immediate suppliers. Encrypting header information would avoid business "ethics" abuses, e.g., stealing one's customer's customer.

Security is also needed to insure that the information collected by third party companies for qualification

_____
1. There were more than 5 million Internet sites by early 1995.

purposes can not be aggregated in other ways to extract proprietary information. For example, if a qualification company saw the full content and timing of a customer's demands and orders, it would be possible to infer the BOM.

6.5 Custom vs. Commodity Products

Implicit in this paper is the view that electronic equipment manufacturing is primarily an instance of buyer oriented commerce, i.e., transactions are initiated by a downstream customer. The desired product may be one that already exists, or it may be a new "custom" product. In either case, the supplier is manufacturing it to meet an actual demand or a forecast demand for that specific item. The supplier is not manufacturing items purely on speculation that customers can subsequently be found.

This is very different from, for example, the soap industry, in which product is manufactured to produce FGI. The assumption here is that marketing will create customers for the FGI. The soap industry is seller oriented, i.e., transactions are initiated by an upstream supplier. The soap manufacturer sells to wholesalers, they sell to retailers, and they sell to end-users. The products are commodities, i.e., the end-user may be satisfied by many equivalent products sold by many equivalent channels.

In reality, electronic equipment manufacturing has both custom and commodity products. At the most upstream end, unrefined silicon is a commodity. Way downstream, personal computers are also commodities, at least for most instances of home use.

When seller oriented commerce predominates, one can imagine a system complementary to DAO, in which the major steps are Availability-Consumptively-Supply. In ACS, a supplier sends availability schedules to potential buyers. They respond with information on the rate at which they would be willing to consume the proffered goods. The supplier then firms up a schedule for which it will commit to supply the goods, and the buyer issues a commitment to take them.

With minor changes, DAO methodology is applicable to ACS. Additionally, DAO is compatible with a range of production scheduling policies, some of which are more applicable to custom products and some to commodity products.

7. Capacity Management

7.1 Manufacturing Resource Capacity

Manufacturers use resources to make products. During each time period, the finite capacity of these resources limits overall production. For example, a card assembly machine may be able to insert electronic components at a peak rate of 2 per second. In an 8 hour shift, therefore, this machine can produce at most 480 cards with 120 components per card. In actual use, it would produce less, because of setup time for the whole job,

load time for each card, and possibly down time due to various failures.

Capacity limits are very difficult to estimate. Setup time, down time, and repair time all tend to reduce capacity. On the other hand, there is often flexibility to increase capacity by using alternate means of production. For example, a machine might be run for 3 shifts rather than 1 shift, or the cards might be vendored out to another supplier, or they might be assembled by hand. Or there might be another machine, committed to a job of lower priority, that could be temporarily "liberated" in an emergency. Within limits, therefore, capacity is a function of how much one is willing to spend for it.

Also, production generally depends on the concurrent availability of different forms of capacity. For example, the assembly of cards would be limited by the simultaneous availability of machines, fixtures for those machines, and operators to run those machines. The capacity of the personnel department might limit production ramp-up and ramp-down capabilities. "Shipping-dock" capacity, mentioned earlier in the context of output shortfall, is another example.

As each plant commits to meet progressively more orders, the capacity of these resources is implicitly allocated. Because of the complexity and sponginess of capacity, however, it is very hard to formalize this allocation in some sort of "capacity schedule". Nevertheless, capacity limits have serious implications.

In the airline industry capacity is simply the number of seats on a flight. It is standard practice for airlines to overbook this capacity because of demand variance; customers fail to honor their commitments. Overbooking places the customer at risk, but the risk is small because there are simple ways to incent people to give up seats. In manufacturing, by contrast, variance may be in supply or demand. In the former case, the risk to the customer may be enormous, and there are no simple solutions.

7.2 Constrained Resource Planning

Constrained Resource Planning (CRP) is the bane of MRP systems. As mentioned earlier, a flaw in simple MRP is that the Master Production Schedule (MPS) it produces may be infeasible because of insufficient capacity. This flaw leads to inflation of PPTs and the consequent unraveling of MRP. Although MRP-II is intended to address this problem, it has not met with widespread success. To avoid similar problems with DAO, therefore, it is essential to explore how CRP might be managed under DAO.

There are three goals for CRP under DAO. The first is to avoid infeasibility, i.e., overbooking. The second is to avoid underbooking, which is costly to the manufacturer. The third is to resolve overbooking, if it does occur, in an unbiased way. For example, other things being equal, there is a "moral" obligation to decommit orders LIFO[1]. In practice, of course, other things would not be equal. Two important differences between MRP and DAO with regard to CRP are in the level of detail and the timing. Fortunately, they match, i.e., for DAO, the CRP timing requirements are very stringent, but the

---
1.   Last In First Out

necessary level of detail is very coarse.

In the MPS generated by MRP or MRP-II, the level of detail is typically that of individual machines or of "sectors" containing a small number of equivalent machines. In a given factory, there may be thousands of such machines or hundreds of such sectors. This is far too detailed a level for supply network production scheduling. For a DAO based system, in most situations, the right level of detail approximates each factory coarsely as one infinitely divisible factory. At any moment, various fractions of this factory capacity are allocated to different assembly orders.

The timing of CRP in MRP is based on periodic batching of orders. Capacity overbooking is not detected until the batch is finally run, at which time the MPS may be found to be infeasible. In DAO, the transactions (demand, availability, order, commitment) overlap one another in real time. If transactions were to be batched, large indeterminate delays would be introduced across the network, and the entire paradigm would collapse. Since the transactions must be processed essentially as they occur in real time, there is no quiescent period during which MPS infeasibility can be corrected.

To avoid infeasibility, therefore, DAO requires some means of performing real time coarse CRP at each supplier site. It is not required, however, that there be any uniformity of these systems across the network.

Invention of a viable local real time coarse CRP system is a research topic beyond the scope of this paper. However, since the viability of DAO, or more generally of any JEI paradigm for supply network production scheduling, is likely to depend on the development of such a real time coarse CRP system, we feel some need to offer a possible solution. In the following sections, therefore, we propose a scheme based on capacity estimation, management, and reservation in each factory. Although this scheme is based solely on intuition and is completely untested, we are confident that some similar approach would be workable.

7.2.1 Capacity Estimation

In an initialization phase, each product j is assigned a number k_j from 0 to 1 that represents Production Engineering's estimate of the relative level of factory capacity needed per unit assembly of that product. In other words, the absolute level of factory capacity needed per unit assembly of that product is αk_j where α is an unknown constant of proportionality.

On any given day we define λ_j is the throughput of product j through the factory's capacity; we define the summation $S = \Sigma \ k_j \lambda_j$ as the computed total capacity load of all products; and we define U as the estimated total capacity utilization (between 0 and 1). If the k_j values are accurate estimates, then there should be a relation of the form $U = \alpha S$ that holds over time, with α remaining constant.

Whenever a demand or order is encountered for x_j units of a product, that product's k_j is determined, and the nominal capacity requirement for that job is a $k_j \ x_j$. The nominal capacity requirement R for the whole factory at any moment of time is the summation $\Sigma \ \alpha k_j x_j$. As the factory is run, the differences between nominal

Constrained Resource Planning

factory capacity load R and estimated factory utilization U are used to drive a regression process that adjusts the k_j values continually, relative to one another and absolutely, to keep α reasonably constant.

7.2.2 Capacity Management

Each supplier dynamically computes a capacity schedule, which is defined to be a list of schedule items of the form {predicted_capacity_requirement, start_date}. (The end date is implicit in the next schedule item in time sequence.) At any instant, this computation is based on the set of currently active DAO schedules. If the supplier wishes, this computation may manage risk by weighting the capacity needs of each DAO schedule and order based on the validity number discussed earlier. Uniform weights of 1 represent minimum risk to customers but present a likelihood that the supplier's capacity will eventually be underutilized. The supplier might also factor in a numerical allowance for the reliability of the customer, as reported by some third party qualification service.

Whenever a new DAO schedule is received, the supplier rationalizes that schedule with its current capacity schedule. The rationalization process delays each DAO schedule item until the earliest time at which the capacity will "suitably accommodate" it. Individual suppliers have considerable freedom in determining what constitutes suitable accommodation. For instance, a supplier might legislate a time-based policy in which its predicted capacity requirement should not exceed 0.25 at a time 6 months in advance, 0.5 at a time 3 months in advance, 0.8 at a time 1 month in advance, and 0.95 at a time one day in advance. Additionally, there might be adjustments based on whether or not the

incremental resource is small (e.g., < 0.01) or large (e.g., > 0.1).

7.2.3 Capacity Reservation

Suppliers might find it useful to institute a capacity reservation system. When an order schedule is received, they could compute the capacity schedule implied by that order in isolation. They could then offer to sell "commensurate" capacity to the customer. For example, suppose an order that calls for 100 items to be delivered in 3 months has an implied capacity schedule that needs 10% of total plant capacity for month 2. The customer, given the opportunity to buy "commensurate" capacity, might then elect to buy 20% of capacity for month 2. (The supplier selling this capacity would be honor bound not to sell more than 100% of capacity during any period; the honor would be upheld by third party qualification companies.) Such a reservation could insure that if there were any unexpected constraints that reduced production by up to 50%, the customer would still get all of its work done during month 2.

8. Information System Architecture

We do not yet have a detailed design for the overall system at each manufacturing site. However, we believe that it will have the general structure shown in Figure 3.

The central block is the DAO supervisor. It maintains a list of activities needed for production scheduling and invokes the schedule processor to handle each one. The schedule processor examines input schedules and derives output schedules. The pricing, capacity, and
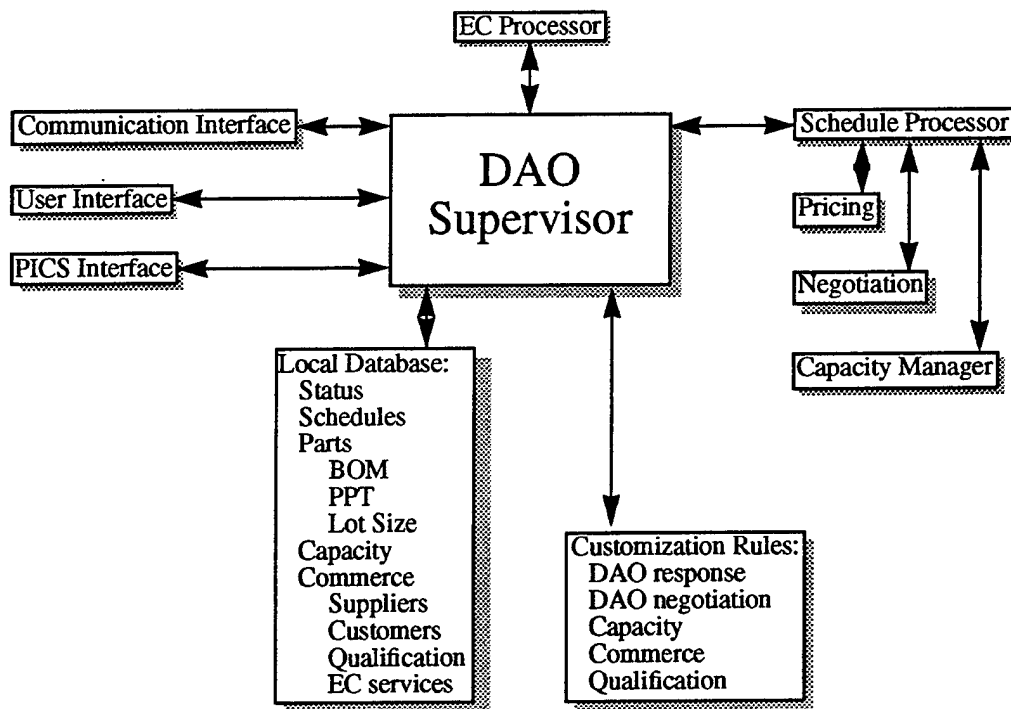

FIGURE 3.   Block structure of overall JEI Supply Network Production Scheduler at each manufacturing site.

Information System Architecture   20 of 23

negotiation modules are invoked by the schedule processor to handle these specialized but complex topics. Actual schedule I/O is performed by the communication interface, which also processes any audit-trail headers. The user interface provides monitoring information to the user and permits supervisory control. The PICS interface and a set of wrappers provide access to a commercial or home-grown Production Information Control System. The EC processor supports DAO with conventional EC activities, such as procurement, billing, and electronic funds transfer (EFT).

The local data base contains several types of information. First, it has data on the content and status of all pending or active schedules. Each time the DAO processor is awakened by the operating system, it inspects this information to revise its to-do list. Second it has data that mirrors information in the PICS, including part design information and WIP inventory data. It also has capacity information that may or may not be related to machine utilization information in the PICS. Finally, it has "commerce" information, including local directories of potential suppliers, customers, providers of qualification services, and other EC services such as EFT.

The rules data base contains customization information dealing with the process of choosing suppliers (or customers) from a list of possibilities, e.g., for determining who to submit demand schedules to. The rules database also has specifications for how to do capacity management, etc.

The schedule processor would apply these customization rules when analyzing schedules and deriving schedules.

Collectively, this system can be viewed as one "agent" in a distributed agent system. The other agents would be similar systems at other manufacturers' sites. For the overall system to operate, there would need to be agreement on communication protocols and data formats in order to exchange schedules, etc.

The system would have the requirements for redundancy and security that are characteristic of any transaction processor. Additionally, it would need some form of security to insure that agents at remote sites were following global rules. We have not yet estimated the overall I/O bandwidth, processing, and storage requirements.

9. Conclusion

The biggest challenge in supply network production scheduling derives from the need to share certain data among autonomous competing companies. This challenge motivates the need for a new production scheduling paradigm based on Just-Enough Information. This paper shows how the DAO algorithm achieves some of the benefits of prior production scheduling methods while avoiding their disbenefits. More importantly, DAO minimizes the need to share proprietary data. Although supply networks have become an essential aspect of modern manufacturing, especially electronic equipment manufacturing, no current production scheduling method offers these features.

In the future world we envision, DAO transactions ripple across the network like rain hitting a pond. Messages flow from one company to another. Within companies, PICS are queried and updated. Deals are negotiated, capacity is reserved, orders are placed, production starts, and parts flow. Ripples also spread as demands change and parts variances trigger alerts. Across thousands of locations, the process repeats hundreds of times an hour, 24 hours a day. Off to the side, qualification service companies monitor the activity, collect data on the performance of suppliers and customers, and make this information available to customers and suppliers.

This paper has presented an initial formulation of a new JEI production scheduling methodology, based on exchange of Demand-Availability-Order schedules, and applicable to manufacturing supply networks in an EC environment. We believe that this new paradigm could result in significant reductions in indirect labor overhead, inventory carrying cost, and excess capacity, at the same time that it increases responsiveness to fluctuations in supply and demand.

[1]   Ammons, J.C., DePuy, G.W., Ellis, K.P., Fisher, M., McGinnis, L.F., Tovey, C.A. Process Optimization for Circuit Card Assembly, *Proceedings of the 1995 NSF Design and Manufacturing Grantees Conference*, January 4-6, 1995, La Jolla, California.

[2]   Annual Survey of Manufacturers. Value of Product Shipments. US Department of Commerce, Bureau of the Census. 1991. [published annually].

[3]   Annual Survey of Manufacturers. Statistics for Industry Groups and Industries. US Department of Commerce, Bureau of the Census. 1991. [published annually].

[4]   Atwater, B. J., Chakravorty, S. S. Does protective capacity assist managers in competing along time-based dimensions. *Production and Inventory Management Journal*, 35(3), 1994. 53-59.

[5]   Baudin, M. *Manufacturing Systems Analysis with Application to Production Scheduling*. New Jersey: Prentice Hall, 1990.

[6]   Brennan, J., Gupta, S. M. A structured analysis of material requirements planning systems under combined demand and supply uncertainty. *International Journal of Production Research*, 1993, 31(7), 1689-1707.

[7]   EM- Electronic Markets. *Newsletter of the Competence Centre Electronic Markets*, University of St. Gallen, Institute for Information Management, Dufourstrasse 50, St. Gallen, Switzerland CH-9000. 1995 [published at regular intervals].

[8]   Neches, A. FAST - A research project in electronic commerce, *EM- Electronic Markets*, October 1993, 9-10, 25-27.

[9]   Friedman, James W. *Game Theory with Applications to Economics*. New York: Oxford University Press, 1990.

[10]   Goldman, S. L., Nagel, R. N., and Preiss, K. *Agile Competitors and Virtual Organizations*. New York: Van Nostrand Reinhold, 1994.

[11]   Goldratt, E. M. *The goal: a process of ongoing improvement*, New York: North River Press, Rev. ed., 1986.

[12]   Hammersley, J. M., and Handscomb, D. C., *Monte Carlo Methods*. London: Methuen & Co., Ltd., 1967.

[13]   Huchzermeier, A., and Cohen, M. A.Valuing Operational Flexibility under Exchange Rate Risk, to appear in *Operations Research, Special Issue on New Directions for Operations Management Research*, 1995.

[14]   Johnson, M. E., Davis, T. Improving Supply Chain Performance Using Order Fulfillment Metrics, *Hewlett-Packard report*, September 1994, Hewlett-Packard Company, 300 Hanover Street, Palo Alto, CA.

[15]   Kekre, S., Mukhopadhyay, T. Impact of electronic data interchange technology on quality improvement and inventory reduction programs: A field survey. *International Journal of Production Economics*, 1992, 28(3), 265-282.

[16]   Miller, G. J., and Vollman, T. E. The Hidden Factory. *Harvard Business Review*, 1985, 63(5), 142-150.

[17]   Monden, Y. *Toyota Production System*. Georgia: Industrial Engineering and Management Press, 1983.

[18]   Mujtaba, S. Enterprise Modeling and Simulation: Complex Dynamic Behavior of a Simple-Model of Manufacturing. *Hewlett-Packard Journal*, December 1994, 45(6), 80-112.

[19]   Orlicky, J. *Material Requirements Planning: The New Way of Life in Production and Inventory Management*. New York: McGraw-Hill, 1975.

[20]   Siha, S. The pull production system: modelling and characteristics. *International Journal of Production Research*, 1994, 32(4), 933-949.

[21]   Silver, E.A., and Peterson, R. *Decision Systems for Inventory Management and Production Planning*. New York: Wiley, 1985.

[22]   Spearman, M. K., Zazanis, M. A. Push and pull production systems: issues and comparisons. *Operations Research*, 1992, 40(3), 521-532.

[23]   Srinivasan, K., Kekre, S., Mukhopadhyay, T. Impact of electronic data interchange technology on jit shipments. *Management Science*, October 1994, 40(10), 1291-1304.

[24]   Sugimiri, Y., Kusunoki, K., Cho, F., and Uchikawa., S. Toyota Production System and Kanban System: materialization of a just-in-time and respect-for-human system. *International Journal of Production Research*, 1977, 15(6), 553-564.

[25]   World Wide Web URL: http://riskweb.bus.utexas.edu/commerce.html

[26]   World Wide Web URL: http://part.net

[27]   World Wide Web URL: http://www.awa.com

[28]   World Wide Web URL. http://www.industry.net

[29]   World Wide Web URL: http://www.internetmci.com

[30]   World Wide Web URL: http://www.marshall.com

[31]   World Wide Web URL: http://www.openmarket.com